Quant Model

Zach Burns

June 10, 2019

```
#{r code = readLines('Code/graphs in paper.R'), echo = FALSE, message = FALSE, warning =
FALSE} #
```

Definitions

Assume the following general case of N task respondents responding to an open-ended prompt, each generating a different number of responses. Let $t \in \{1, 2, ..., N\}$ be a task respondent, each of whom produces j_t *jots*. Thus, $R_{t,j}$ is the j^^th jot produced by the t^^th task respondent. R is the set of all responses $\{r_{1,1}, r_{1,2}, ..., r_{1,j_1}, r_{2,1}, ..., r_{N,j_N}\}$. The count of this set we call the *total mass* of responses. We further define *frequency* as the number of times a jot appears in R. (Possibly can leave out the foolowing sentence) We use *jot* in contexts where the relevant level of analysis is the participant (e.g. a task respondent producing jots) and *response* in contexts where the relative level of analysis is the task (e.g. responses are scored relative to one another in the response set).

For example, imagine a task respondent is given 3 equally-sized LEGO bricks, A, B, and C that can be placed in positions 1, 2, and 3. An observer could track the positions of all bricks, and one possible jot is "BC.0.A" indicating that block B is stacked on top of block C in position 1, position 2 is empty, and block A is in position 3. In this task, there are 99 possible non-empty configurations (Garrett: I did calculate this!), and researchers might be interested in how unique each of the configurations are.

(In the example of a DT task, jots would be individual responses. More generally, jots can be any individuated state or response. In the example of a game with descrete game states, jots would be those states.)

We propose a measure of uniqueness called Mass-Adjusted Uniqueness Index (MAUI). MAUI is a response-level scaling index that defines uniqueness as a function of item and person parameters. (Garrett: I'm not sure exactly how to make this fit, but I'm sure we'll get after it at some point.) Specifically, MAUI defines uniqueness of a given response as the proportion of responses in the sample that were as- or less-frequently. MAUI has the following properties:

1) Within a given response set R, all responses that have the same frequency have equal MAUI scores
2) For a given response $r_{t,j}$ within response set R, all responses that have a lower frequency have lower MAUI scores
3) The score is bounded between 0 and 1.
4) The average of MAUI scores of all responses in a sample is always 1/2 (see proof in appendix).

Example

Suppose the following makeup of responses in a hypothetical sample:
Responses A_1, A_2, A_3, and A_4 were each given 5 times.
Responses B_1, B_2, and B_3 were each given 4 times.
Responses C_1 and C_2 were each given 3 times.
Responses D_1 and D_2 were each given twice.
Responses E_1 through E_8 were each given once.

The responses with the same letter were given an equal number of times, and thus will have the same MAUI score. The mass of responses A_n is 20, B_n is 12, C_n is 6, D_n is 4, and E_n is 8, for a total response set mass of 50. Each response is given a MAUI score:

Responses A_n each receive a score of $\frac{1}{2} * \frac{20}{50} = .2$.
Responses B_n each receive a score of $\frac{20}{50} + \frac{1}{2} * \frac{12}{50} = .52$.
Responses C_n each receive a score of $\frac{20}{50} + \frac{12}{50} + \frac{1}{2} * \frac{6}{50} = .7$.
Responses D_n each receive a score of $\frac{20}{50} + \frac{12}{50} + \frac{6}{50} + \frac{1}{2} * \frac{4}{50} = .8$.
Responses E_n each receive a score of $\frac{20}{50} + \frac{12}{50} + \frac{6}{50} + \frac{4}{50} + \frac{1}{2} * \frac{8}{50} = .92$.

The average of all MAUI scores for the 50 responses is:

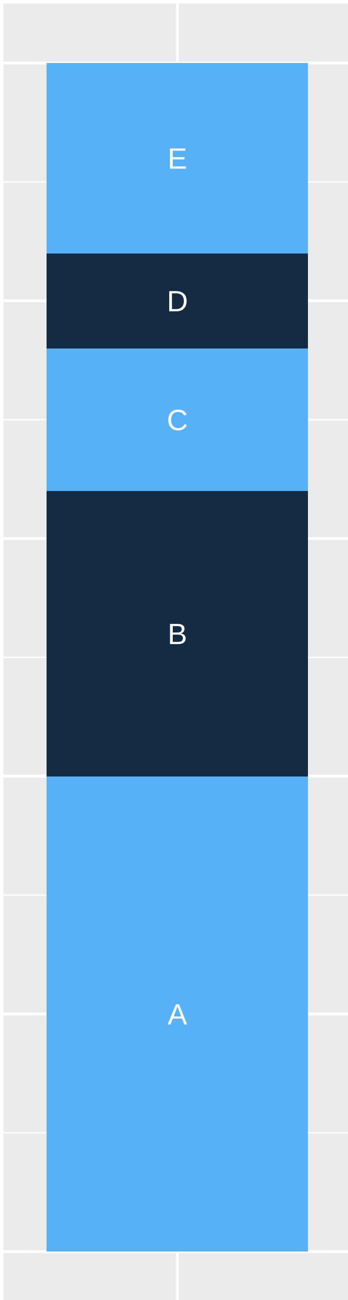
$$\frac{20}{50} * .2 + \frac{12}{50} * .52 + \frac{6}{50} * .7 + \frac{4}{50} * .8 + \frac{8}{50} * .92$$

$$= .08 + .1248 + .084 + .064 + .1472 = .5$$

Visualizing Uniqueness

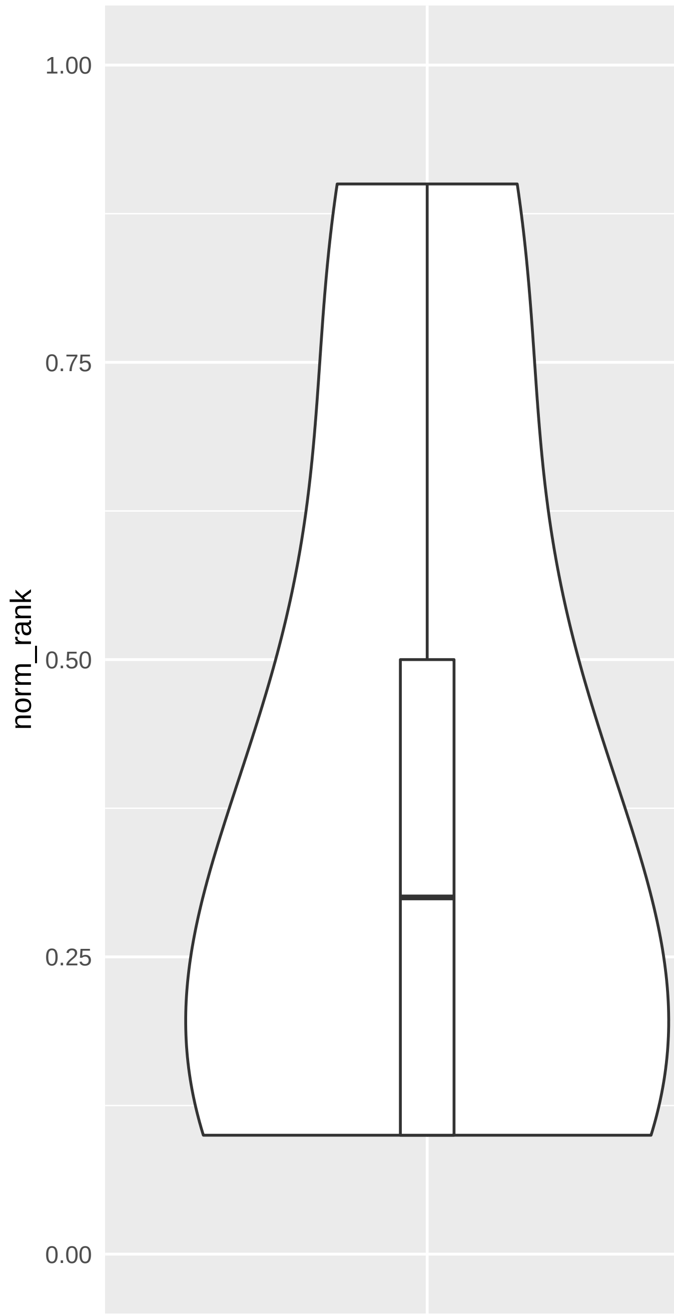
We return to the metaphor of “low-hanging fruit” in order to visualize MAUI scores. In the metaphor, lower fruit are easier to reach, and therefore more common (and less unique).

Using the above example, all of the A_n responses are the lowest-hanging fruit. Because they are all equally frequent in the sample, they have equal MAUI scores, and are indistiguishable in uniqueness. All together, the 4 A_n responses account for 40% of the sample.



In the preceding plot, each color grouping represents the mass of responses that were given an equal number of times. The center of each mass (i.e. where each group is labeled) represents the MAUI score of that group, as a percentage of the whole bar.

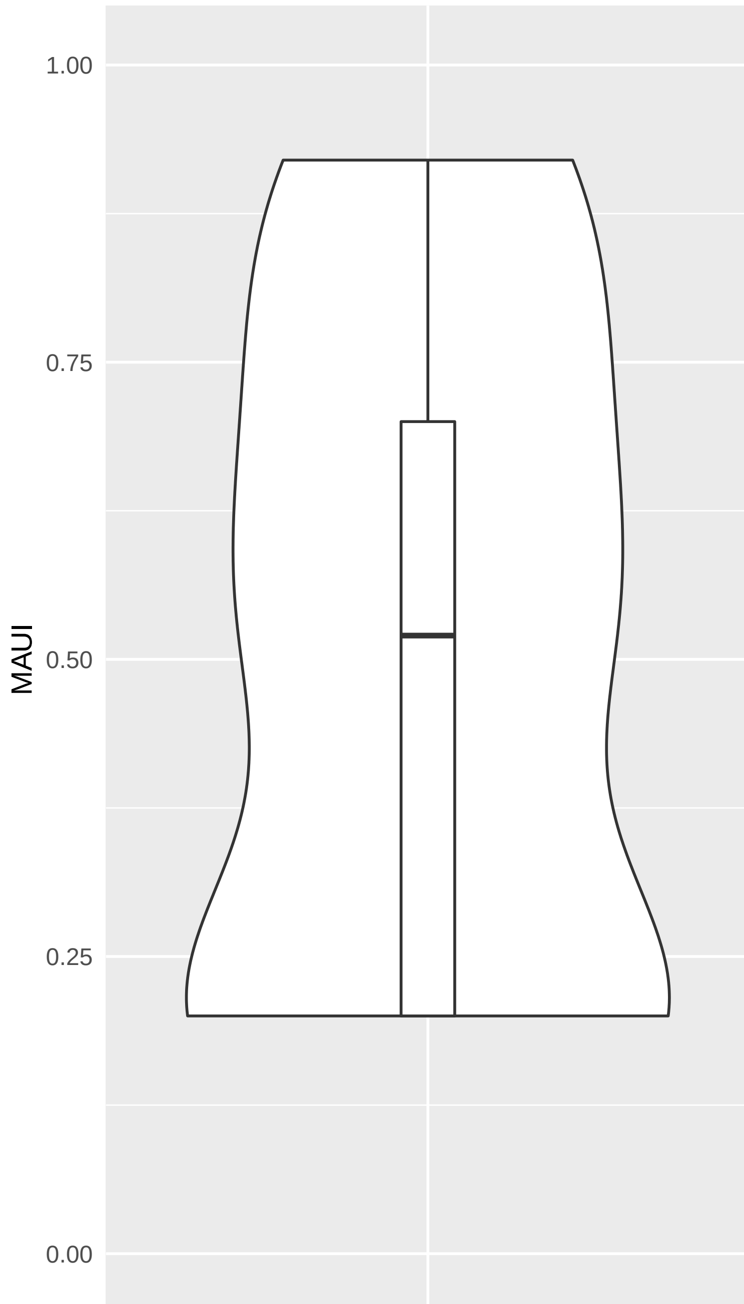
If we imagine all response frequency groups hanging equidistantly in the ideation tree, and plot a smoothed his-



togram with a box-and-whiskers in the middle, we get the following:

In this plot, response frequency groups A, B, C, D, and E are graphed at .1, .3, .5, .7, and .9, respectively (i.e. equidistant and normalized). This plot provides a visual interpretation of where the masses of responses are. In the example, there are more responses that were frequently given, creating a fatter bottom and bringing the median (black line in the box-and-whiskers plot) to roughly 1/3.

we create a similar graph showing the MAUI scores across the entire sample. That is, instead of imagining the response frequency groups hanging equidistantly, we imagine they are at the height given by their MAUI



score:

With the MAUI score on the y-axis, and the frequency of responses on the x-axis, his plot essentially shows a smoothed histogram of MAUI scores from the example, with a box-and-whiskers plot in the center. (Note that the black line is the median value, not the mean, which would be at exactly .5) In actual samples with a greater number of responses, the sides of the plot will be nearly straight. MAUI was designed to have exactly this property.

Proof that weighted average of MAUI scores is always 1/2.

Let m_i be the frequency of the i$^{\text{th}}$ most frequently given answer.

Let $M_n = \sum_{i=1}^{n} m_i$; that is, the sum of all response mass frequencies from 1 (the largest) to n. If n is the rank of the least frequently given response (which will generally be those given by a single participant), then M_n is the total count of all jots given by all participants in the sample.

Therefore:

$$M_n^2 = (\sum_{i=1}^{n} m_i)^2$$

$$= (\sum_{i=1}^{n} m_i) * (\sum_{j=1}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i) * (\sum_{j=1}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i) * (m_i + \sum_{j \neq i}^{n} m_j)$$

$$= (\sum_{i=1}^{n} m_i^2) + (\sum_{j \neq i}^{n} m_i m_j)$$

$$= (\sum_{i=1}^{n} m_i^2) + 2(\sum_{j<i}^{n} m_i m_j)$$

MAUI is defined as the mass of all responses given more frequently plus the center of the mass of responses given equally frequently. The weighted average of MAUI scores is therefore defined as:

$$\overline{MAUI}_n = \sum_{i=1}^{n} (\frac{m_i}{M_n})(\frac{\frac{m_i}{2} + \sum_{j=0}^{i-1} m_j}{M_n})$$

$$= \sum_{i=1}^{n} (\frac{m_i}{M_n})(\frac{m_i + 2\sum_{j=0}^{i-1} m_j}{2M_n})$$

$$= \frac{1}{2M_n^2}[\sum_{i=1}^{n} (m_i)(m_i + 2\sum_{j=0}^{i-1} m_j)]$$

$$= \frac{1}{2M_n^2}[\sum_{i=1}^{n} ({m_i}^2) + 2(\sum_{i=1}^{n}\sum_{j=0}^{i-1} m_i m_j)]$$

$$= \frac{1}{2M_n^2}[\sum_{i=1}^{n} ({m_i}^2) + 2(\sum_{j<i}^{n} m_i m_j)]$$

As shown above,

$$(\sum_{i=1}^{n} m_i^2) + 2(\sum_{j<i}^{n} m_i m_j) = {M_n}^2$$

The resulting quantity is 1/2 for all n. $\square$